Similarity analysis for the Blue Nile basin, in the Ethiopian Highlands

By Catherine Pfeifer, An Notenbaert, Abisalom Omolo

1 Introduction

Like in large parts of sub-Saharan Africa, most farmers in the Ethiopian highlands base their livelihood on unreliable rain fed agriculture. Climate variability as well as the lack of water management explains to a large extend the prevailing food insecurity and poverty (Hanjra & Gichuki 2008; De Fraiture et al. 2010). Consequently, improving rainwater productivity is an appealing solution to alleviate hunger in the region (Hanjra et al. 2009; Hanjra & Gichuki 2008).

Integrated rainwater management is a recently developed concept that abandons the differentiation between irrigated and rain fed agriculture (Humphreys et al. 2008; Rockström et al. 2010; Rockström et al. 2003). It encompasses any bundle of practices, referred to as a strategy that aims at increasing rainwater productivity. Rainwater management strategies (RMS) include practices such as increasing soil water holding capacity, enhancing crop and livestock water productivity, improving efficiency of small scale irrigation, efficient use of ground water wells, diversion, or water harvesting (Johnston & McCartney 2010; Hanjra & Gichuki 2008; Rockström et al. 2010).

RMS are relatively low cost and can potentially be made available to many farmers and communities. Despite of all these benefits, adoption rates of RMS practices are low and dis-adoption is important. Indeed, up until today, RMS practices have been promoted regardless of site-specific bio-physical characteristics and regardless of the socio-economic and institutional environment (Faures & Santini 2008). For a more successful promotion of RMS, a paradigm change towards promotion of location-specific interventions is needed (Faures & Santini 2008). Beyond bio-physical suitability, successful implementation crucially depends on farmers' willingness to adoption a practice. Therefore, the socio-economic and institutional environment must be taken into account in a spatially explicit way. A first step towards the promotion of site-specific RMS requires an understanding which sites present similar bio-physical, socio-economic and institutional characteristics within a basin.

The objective of this report is therefore to develop a methodology that allows identifying locations within a landscape that have similar bio-physical, infrastructure, socio-economics, and institutional characteristics relevant to RMS. This methodology is applied to the Blue Nile basin in the Ethiopian highlands.

2 Methodology

2.1 Study area

The Nile Basin Development Challenges program (NBDC) aims at targeting RMS in the Blue Nile basin in the Ethiopian Highlands. Within the Blue Nile basin, specific study landscapes representing dominant farming systems in the Ethiopian Highlands will be the basis for research. Smaller action research sites

(ARS) are further embedded within these study landscapes providing a nested set of sites for learning and research at a variety of physical and social scales. Based on expert consultation and discussion within the NBDC project team, the following criteria were used for selecting these study landscapes:

- *Represent diversity* found in the Ethiopian Blue Nile in terms of:
    - Agro-ecologies and livelihood strategies including important cropping systems, perennial and agropastoral systems
    - Water availability
    - Socio-economic status
    - RWM challenges
    - Market access
    - Diversity of actors
    - State of transition – desperate or blooming?
- Synergy and complementarity with other initiatives, building on not duplicating
- Ability to demonstrate results
- Presence of partners to work with
- Accessibility for at least one site to serve as CPWF show case
- Containing examples of both SLM and AGP Woredas

Based on these criteria, three study landscapes, differing in state of development, agro-ecology (specifically water availability), important livelihood systems and opportunities for RMS were identified: Jeldu, Diga and Fogera (Figure 1)

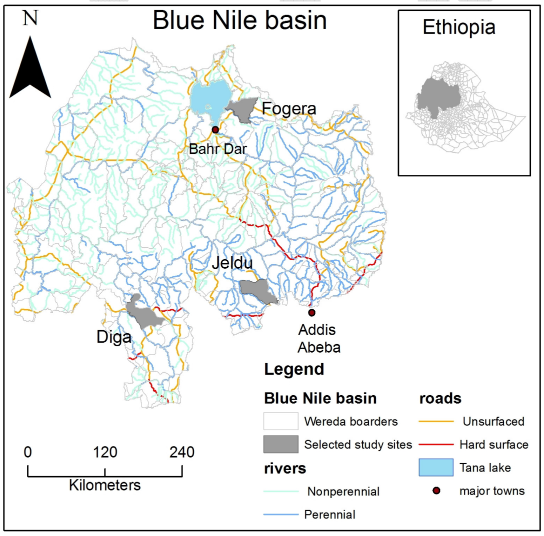

Figure 1 : The Blue Nile basin in the Ethiopian highlands and the selected study sites : Jeldu, Diga and Fogera

2.2 The overall approach

In order to identify similar locations within the Blue Nile basin, two types of similarity analysis can be performed as shown in Figure 1: an analysis of individual characteristics for which data are available and

an analysis of a set of characteristics requiring an aggregation from individual characteristics.

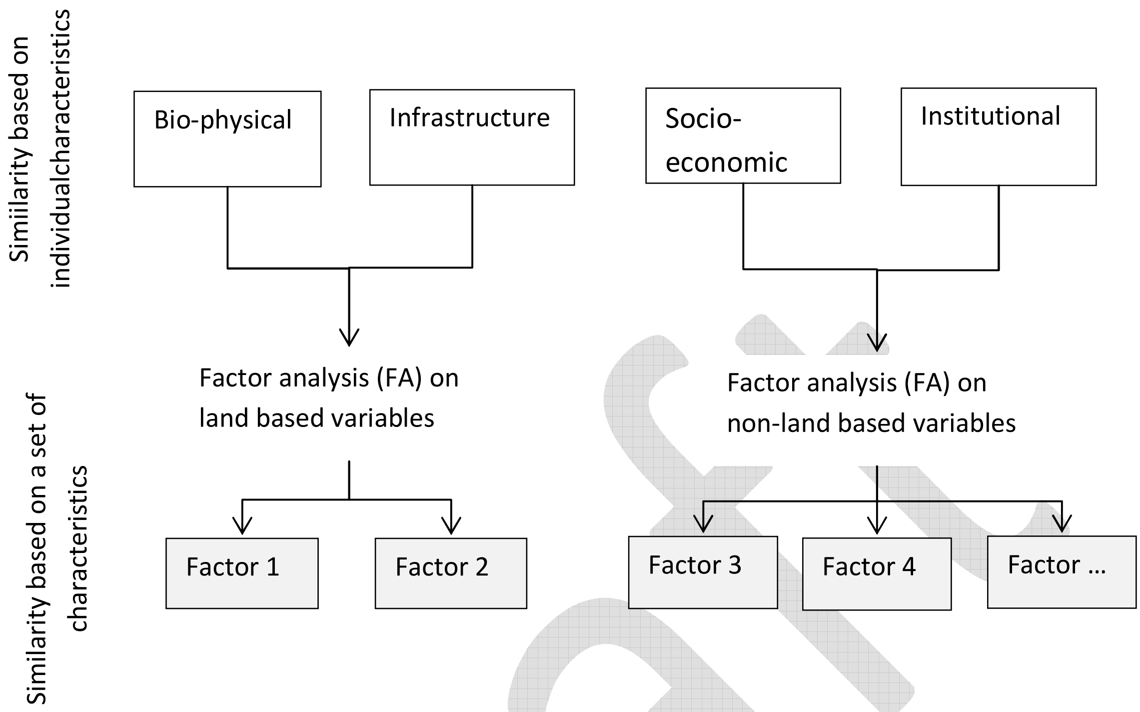

Figure 2 : link between the similarity analysis on the individual characteristics and the similarity analysis on the aggregated characteristics

In a first step, potential drivers of successful adoption for RMS practices need to be identified. Beyond bio-physical drivers that are relatively well understood, literature identifies human dimensions, namely socio-economic and institutional environment as well as infrastructure as potential drivers for adoption of RMS (Baguma et al. 2010; De Fraiture et al. 2010). Therefore individual characteristics can be classified into four categories: bio-physical, infrastructure, socio-economic and institutional, reflecting different disciplines, environmental science, landscape planning, economics/sociology and policy respectively. Because the socio-economic processes adjust much faster than environmental processes and infrastructure, these categories can be classified into land-based processes and non-land based process, reflecting a similar time scale at which the variables change.

In order to aggregate these individual characteristics, a factor analysis is used as data reduction method. This approach allows to identify underlying structures, also referred to as unobserved structure and reduces the individual characteristics into a data driven number of factors. Given that land based and non-land based variables change at different time scales, it does not make sense to assume a similar unobserved construct for both types of variables. Therefore, FAs on land-based and on non-land based data are run separately.

2.3 The similarity analysis on the individual characteristics

Geo-referenced data or proxy data, covering the whole Ethiopian side of the Nile basin, need to be found for each potential driver of success. For the similarity analysis, the data for each characteristic, from now on referred to as variable, needs to be classified. Sites can then be said similar when they fall into the same class. For a relevant similarity analysis, the selection of the right classification technique is therefore crucial. Table 1 shows different classification techniques for different types of data. For some of the variables, classification is not an issue, as the data is collected in a finite numbers of categories. Classification becomes more difficult with data that has too many categories to be visualized or can take any value. In this case one can rely on expert knowledge to define the classes. When no expert knowledge is available then different classification techniques can be applied: equal interval, quantile and natural breaks. The equal interval cuts the distribution of the data at equal breaks and fails to reveal the distribution of the data. Quantile classification cuts the population in equal group size and therefore allows to visualize the distribution of the data. However, gaps between classes can be important when the data contains outliers. The Jenk's natural break technique is a data driven approach. It seeks to reduce the variance within a class and maximize the variance between the classes. This approach depends on the a priori defined number of classes, but when these are selected carefully this classification techniques identifies "depressions" in the distribution of the data and groups together data with similar characteristics. It is therefore a promising technique for a similarity analysis.

Table 1 : classification technique for different types of data

Type of data	Grouping technique	Examples
Categorical data with a finite number options	No grouping	Dominant soil Types of roads
Data with expert knowledge over range	Group into categories based on expert knowledge	High and low soil degradation
Data without any expert knowledge	Equal interval Quantiles Natural breaks	Income

2.4 Similarity analysis based on a set of characteristics

Similarity analysis based on a set of characteristics aims at identifying dimensions within which locations differ and calls for an aggregation of the various characteristics. But environmental and social processes in a landscape are complex and interlinked (Burel & Baudry 2004), and therefore the variables chosen for assessing these processes are likely to be correlated. Factor analysis (FA) is a tool that allows data reduction. It describes variability among observed variables in terms of a potentially lower number of unobserved variables called factors. Factor analysis searches for such joint variations in response to

unobserved latent variables, also referred to as an unobserved construct. The observed variables are modeled as linear combinations of the potential factors, plus "error" terms. The information gained about the interdependencies between observed variables can be used later to reduce the set of variables in a dataset.

Results from a FA are: the factor scores, the transformed variable values and the loadings, the weight by which each standardized original variable should be multiplied to get the factor score. Usually the loadings are the result of a geometrical rotation to minimize the variability within the factor and maximize variability between factors.

Factor analysis and principal component analysis (PCA) though related and often giving very similar results, are not the same. FA estimates how much of the variability is due to common unobserved factors while PCA that takes all variability of the variables into account. When the unexplained variability ("error") of all the variables is equal, FA is a particular case of PCA and both approaches yield to the same result. Beyond data reduction, FA has the advantage to identify an underlying structure, which is particularly interesting for identifying socio-economic dynamics.

The application of FA to spatial data, raises the issue of the spatial scale at which the analysis is run. Whereas bio-physical data are often detailed and available at high resolutions, socio-economic data is often available at administrative level. One strategy is to aggregate all the data to the smallest common spatial unit of the data. For Ethiopia, this implies the woreda level (districts) for which socio-economic data is available. As a consequence, a certain degree of spatial variability of bio-physical variables is lost when aggregated to the woreda level. Standard deviation, maximum and minimum values can be presented as indicators for the lost variability.

The FA on woreda level results in loadings defining each factor. These loadings can be used to predict the factor scores for each woreda, which can then be mapped. The natural break classification technique enables a data driven classification of locations that are similar.

3 Results for individual characteristics

3.1 Selection of relevant characteristics

A comprehensive review of RMS in Ethiopia identifies a whole set of bio-physical, infrastructure, socio-economic and institutional characteristics as drivers of adoption of RMS (Merrey & Gebreselassie 2010). This list has been expanded with expert knowledge collected from ILRI and IWMI experts. Many geo-referenced variable or proxy variable for these drivers could be found. Table 2 shows the chosen variables as well as their source.

Table 2 : Chosen variables, sources and grouping techniques for similarity analysis

Variables

Bio-physical	Source	Classification technique
Rainfall	Jones and Thornton (2000)	Expert
CV rainfall	Jones and Thornton (2000)	Expert
Rivers	FAO geonetwork	None
Wetlands	FAO geonetwork	None
Elevation	SRTM*	Expert/FAO
Slope	SRTM	Expert/FAO
Soil type	Harmonized world soil database FAO	None
Soil texture	Harmonized world soil database FAO	None
Soil drainage	Harmonized world soil database FAO	None
Land degradation	Bai et al.(2008)	None
Temperature Mean	Worldclim	Expert
Minimum temperature °C	Worldclim	Expert
Maximum temperature °C	Worldclim	Expert
Length of growing period	Jones and Thornton (2000)	Expert
Malaria prevalence	MARA/ARMA (Snow et al. 2005)	Expert
Infrastructure		
Roads	ESRI*	None
Travelling time to major cities	IFPRI (Schmidt & Kedir 2009)	Expert
Density of primary school (number of schools/population)	IFPRI REA*	Quantile/ 5 classes
Density secondary schools	IFPRI REA	Quantile/ 5 classes
Agricultural production		
Livestock production systems	Thornton et al (2002)	None
Livestock densities	Gridded livestock of the world, FAO	Expert
Major crops yield	IFPRI REA	Equal interval
Crop utilization (share of crop that is sold on market)	IFPRI REA	Equal interval
Socio-economic		
Percentage of population living below the poverty line	CSI*	Equal interval
Population density	IFPRI REA	Equal interval

Proportion of female headed agricultural households	IFPRI REA	Equal interval
Literacy level for agricultural population aged 15 to 30	IFPRI REA	Quantile/ 5 classes
Proportion of Household Heads Fully Engaged in Agricultural Activity	IFPRI REA	Quantile/ 5 classes
Percentage of holders utilizing hired labor	IFPRI REA	Quantile/ 5 classes
Household size	IFPRI REA	Quantile/ 5 classes
Institutional		
Percentage of holders utilizing credit services	IFPRI REA	Quantile/ 5 classes
Percentage of holders utilizing advisory services	IFPRI REA	Quantile/ 5 classes
Holding fragmentation - average number of parcels per holdings	IFPRI REA	Quantile/ 5 classes
Proportion of land area under smallholders that is rented	IFPRI REA	Quantile/ 5 classes
Percent of crop holders who have land holding size of 1 hectare or less	IFPRI REA	Quantile/ 5 classes

CSI : http://www.cgiar-csi.org/data
ESRI : http://www.esri.com/data/free-data/index.html
FAO geonetwork : http://www.fao.org/geonetwork/srv/en/main.home
MARA/ARMA : http://www.mara.org.za/
IFPRI REA = rural economy atlas (2006) : http://www.ifpri.org/node/3763
SRTM : http://srtm.csi.cgiar.org/
Worldclim : http://www.worldclim.org/download

A number of characteristics were mentioned by experts for which no geo-referenced data could be found. These characteristics included: a detail land cover map, areas under land certification, proximity to veterinary support.

3.2 Maps of individual characteristics

For the purpose of NBDC project, each of the maps indicate the 3 study site woredas, Fogera (North), Diga (West) and Jeldu (East)

3.2.1 Rainfall

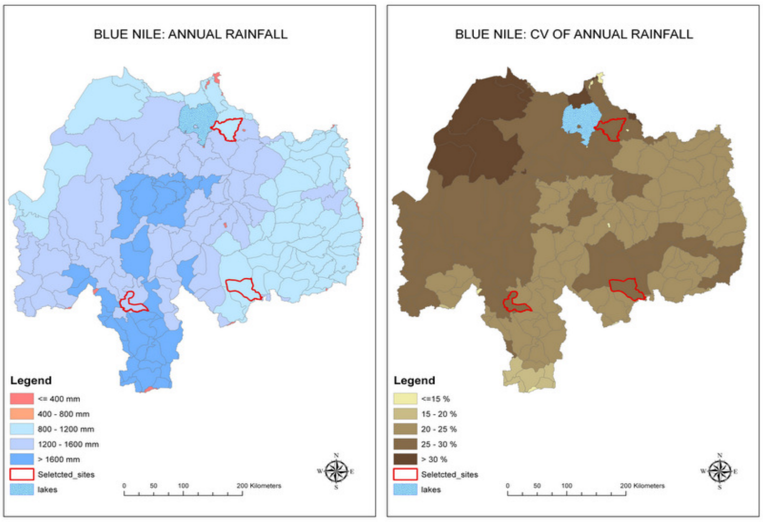

Figure 3 : Rainfall (left) and coefficient of variation of rainfall (right) in the Blue Nile basin with the 3 woredas: Fogera, Jeldu and Diga

Figure 2 shows rainfall and its coefficient of variation for the Blue Nile basin. Locations with higher rainfall have the tendency to have less variable rainfall. Jeldu and Fogera have a similar rainfall patterns while Diga has slightly higher rainfall. All three study sites have the same rainfall variability.

3.2.2 Rivers, wetlands

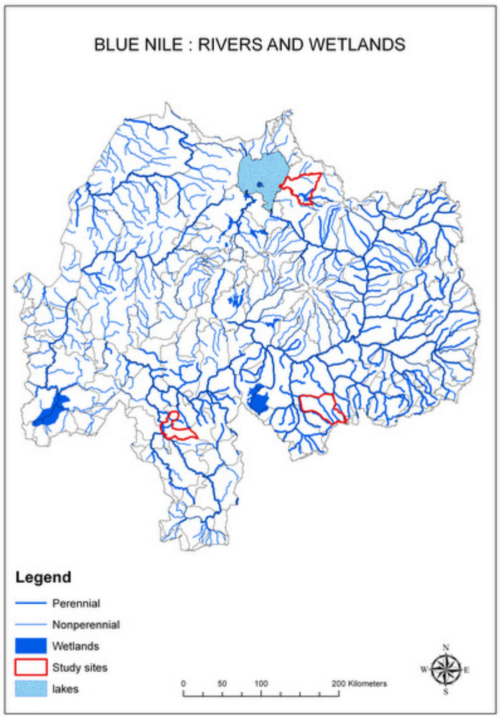

Figure 4 : rivers and wetlands in the Blue Nile basin

Figure 4 shows the rivers and wetland in the basin. It shows that all four of the study sites is crossed or boarded at least by one perennial river. In addition the study site Fogera shores with lake Tana. Furthermore there are no wetlands is the chosen study sites.

3.2.3 Elevation and slope

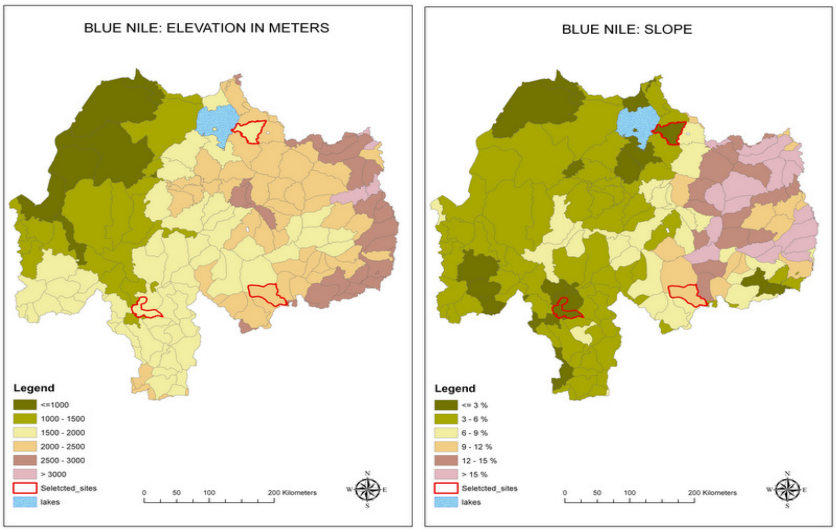

Figure 5 : Elevation and slope in the Blue Nile basin

Topography in the Blue Nile basin follows a gradient from the flat lowlands in the West to mountainous areas in the East as shown in Figure 5. A flat highland plateau crosses the basin from the Tana lake in the North to the South. Diga and Fogera have a similar elevation and slope, while Diga is slightly higher and hillier than the other two sites.

3.2.4 Soils

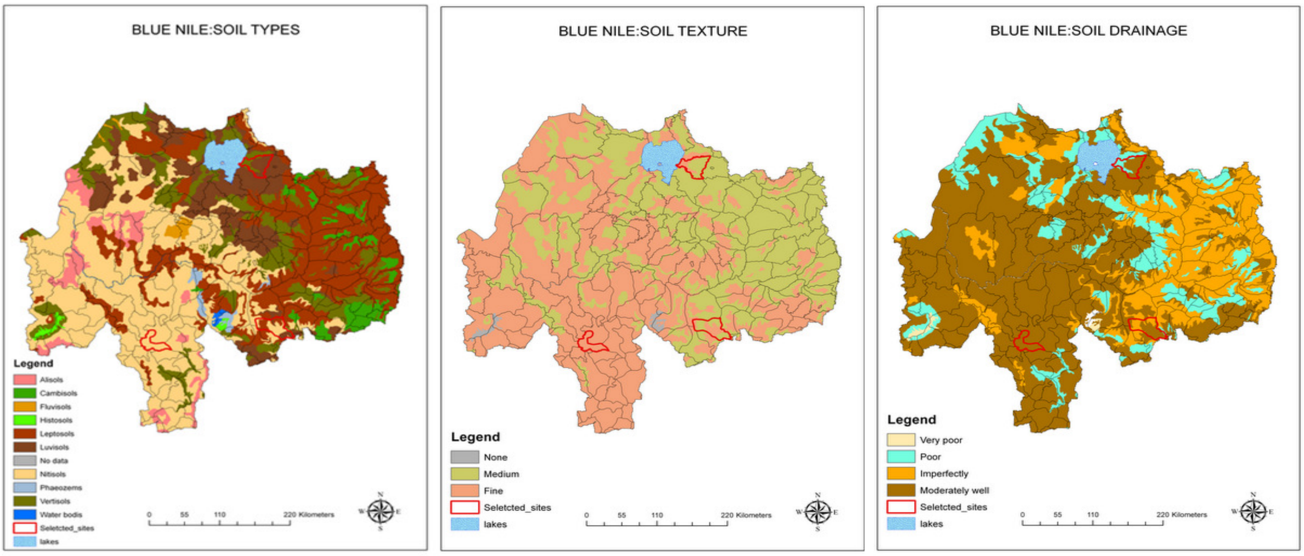

Figure 6 : Soil type (left) soil texture (middle) and soil drainage (right) in the Nile basin. drainage and texture

Figure 6 shows soil type, soil texture and soil drainage. Each soil type has its own characteristics, and there for another management and hydrology. Soil texture indicates how easily water can infiltrate, while water drainage is a combination of soil textures and slopes. Fine soil textures in flat area are likely to be less drained than less fine structure on slopes.

As shown in the left of Figure 6, the western Ethiopian Highlands are dominated by nitisols, the so called “red tropical soil”. Its stable porous soil structure permits deep rooting of plant and makes it less prone to erosion. Internal water drainage, water holding capacity and workability are good. It is a soil that is moderately to highly productive under a wide range of crops (Driessen & Dudal 1991). The East is dominated by leptosols. These soils are relatively shallow and prone to erosion. Consequently, these soils are unattractive for agriculture and have a limited potential for tree production and extensive grazing (Driessen & Dudal 1991). On the highland plateau vertisols can be found. Vertisols are heavy clay soils. They generally have a fine structure and poor internal drainage. Because of these characteristics, workability of the soil is low. Soils need to be well managed in order to be suitable for agriculture. Therefore vertisol plains lend themselves better to mechanized agriculture than for low technology agriculture. In contrast, on vertisols on slope, contour bunding improves significantly productivity the agriculture. Also around Lake Tana extensive Luvisols can be found. These soils are fertile and suitable for a wide range of agricultural uses. The selected study sites are rather different in terms of soils. Jeldu consists mainly of fine relatively well drained leptosols and moderately fine and imperfectly drained nitisols, while Diga is dominated by fine well drained nitisols. Fogera is dominated by relatively well drained and fertile Luvisols and poorly drained vertisols directly adjacent to Lake Tana.

3.2.5 Land degradation

Data behind the land degradation layer is the so called restrend maps from ISRIC (Bai et al. 2008). It makes use of the error between the observed normalized difference vegetation index (NDVI) with a predicted NDVI computed based on the NDVI correlation with rainfall. The analysis of the trend of the error, indicated locations where there is human induced land degradation. Human induced land degradation takes mainly part in the areas with the erosion prone leptosol, around Tana lake as shown in Figure 7.

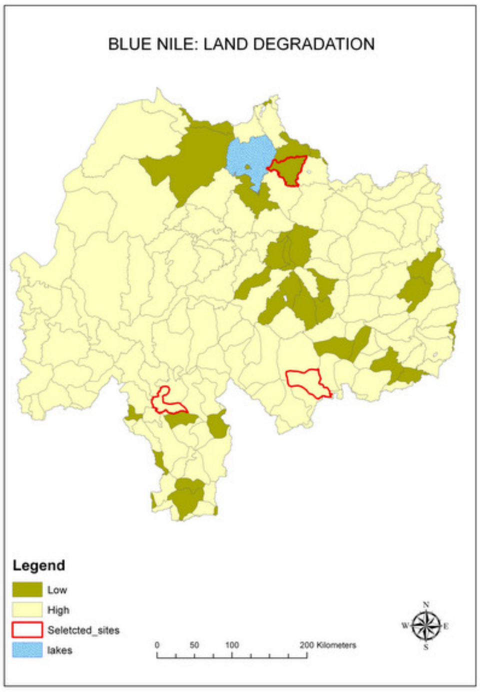

Figure 7: Land degradation in the Nile basin

3.2.6 Temperature

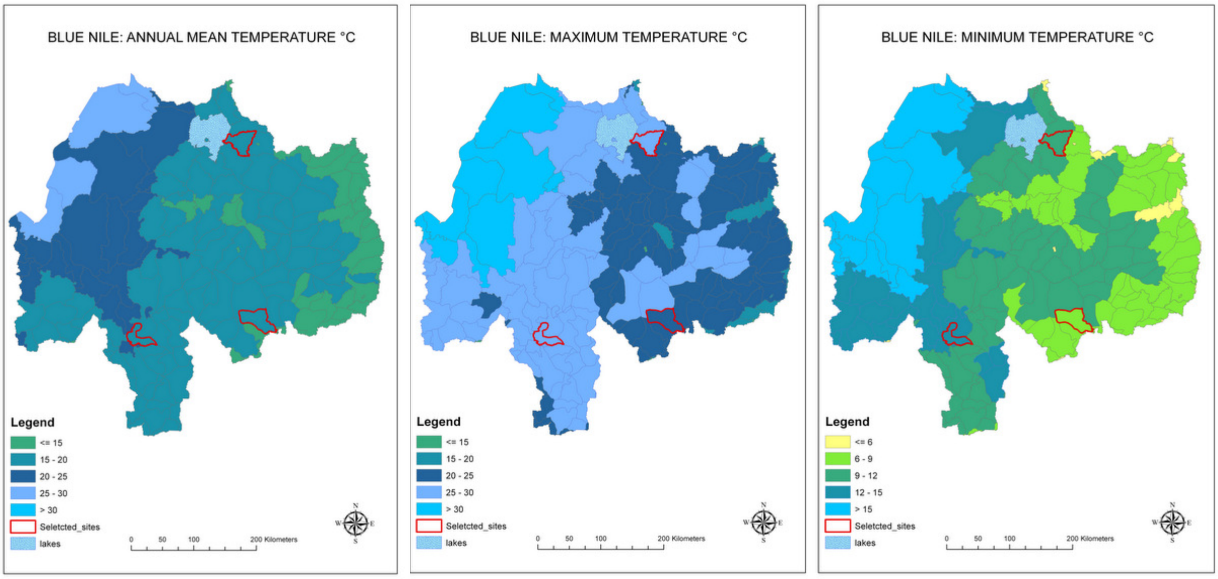

Figure 8 : Annual mean temperature (left), maximum temperatures (middle) and minimal temperature (right)

Annual mean temperature, maximum and minimum temperature show approximately the same spatial pattern than elevation (Figure 8). All three study sites have similar annual temperature, while minimum and maximum temperatures do show differences with Jeldu being the coolest area, and Diga the warmest.

3.2.7 Agri-ecology

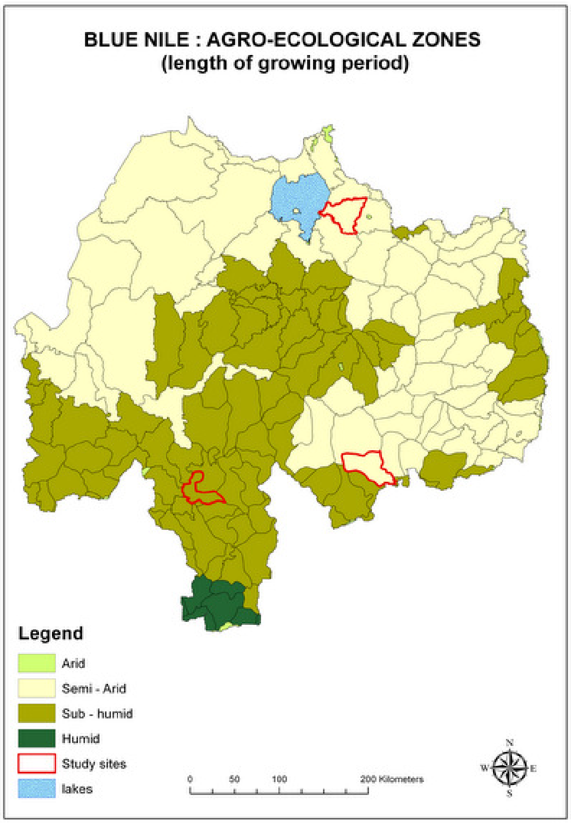

Figure 9 : agro ecological zones based on the length of growing period

Agro-ecological zones map combines the patterns of rainfall and temperature (Figure 9). The humid highlands have the longest growing period and include Diga. Jeldu and Fogera are in the semi-arid areas and have a shorter length of growing period.

3.2.8 Malaria prevalence

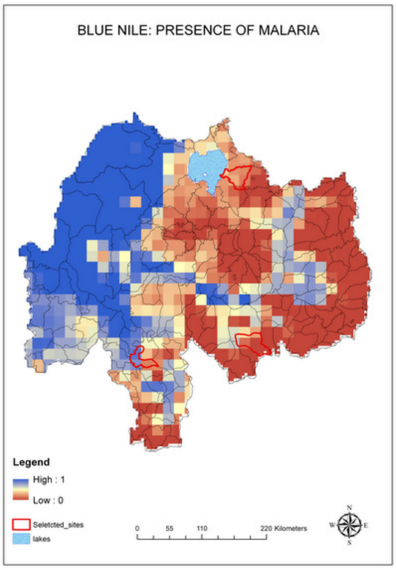

Figure 10 : prevalence of Malaria in the Blue Nile Basin

Malaria mainly prevails in the lowland in the West as well as the in the valleys in the West. All study areas are only moderately affected by Malaria. Diga being located in the lower and warmer locations has the highest probability of malaria.

3.2.9 Remoteness

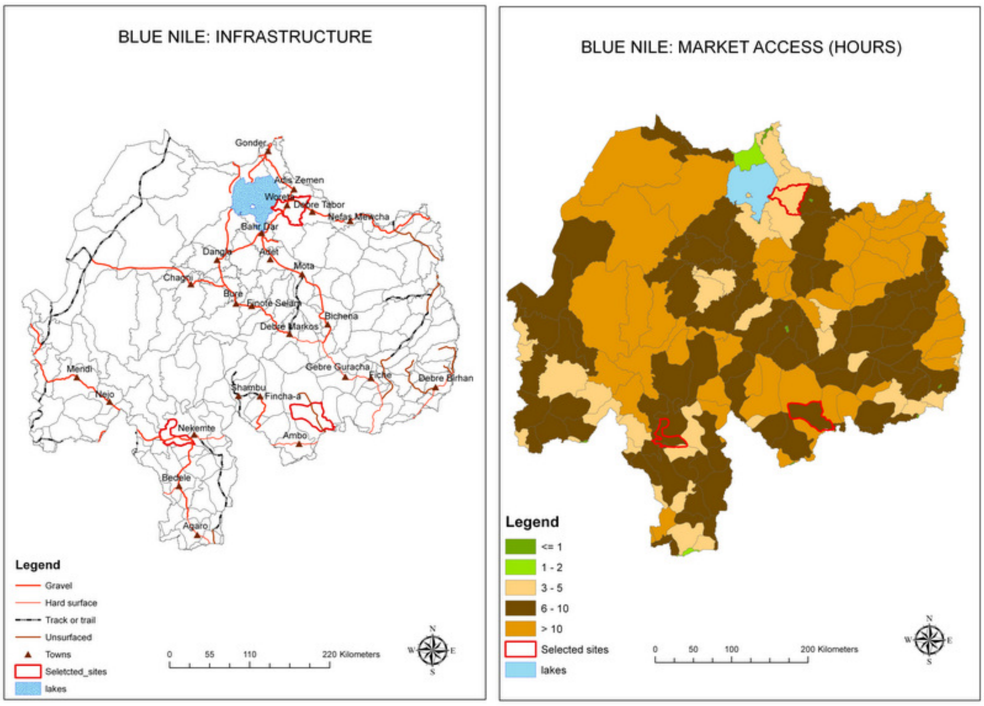

Figure 11 : Road map (on the left) and travelling time to major market/towns (on the right)

Most of the roads link Addis to the Lake Tana region, as well to the West. Fogera is the study site with the best market access, while Diga and Jeldu are less accessible. Nevertheless, this comparison should be taken with caution as new surfaced roads are being built at the time of this study making Jeldu more accessible already today as well as Diga in the near future.

3.2.10 School density

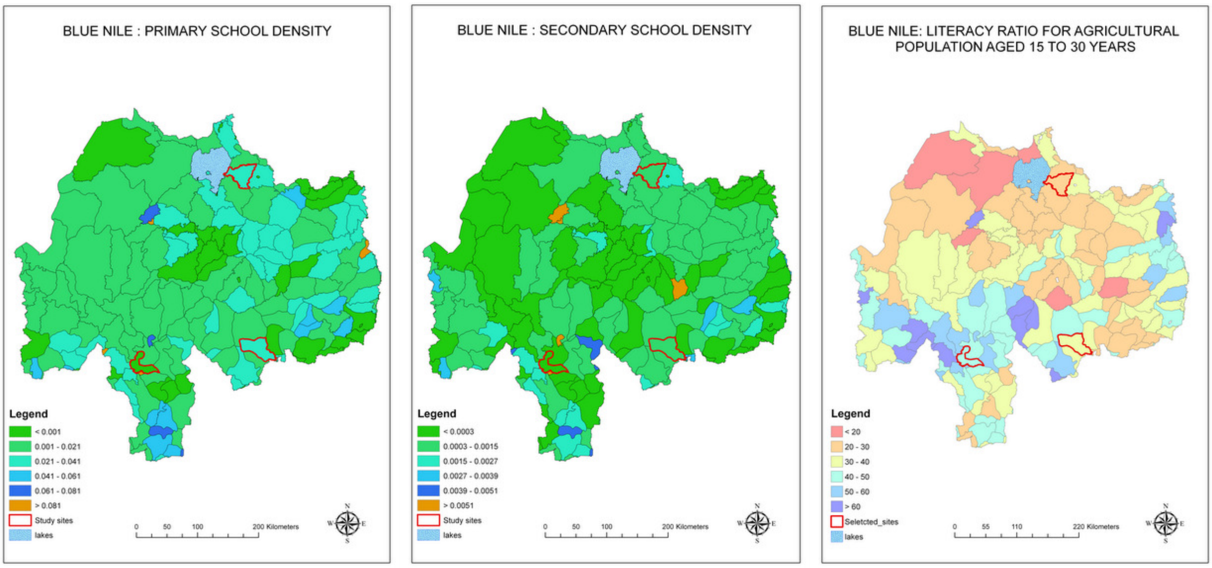

Figure 12 : Primary (left) and secondary (middle) school density, and literacy ratio

Figure 12 shows the school density (number of school/area). First of all, the density of primary schools is higher than the density of secondary schools. Also the East and the South have a higher school density than the West. Interestingly the density of schools does not seem to be correlated with literacy. The South seems to be more literate than the rest of the area. Jeldu and Fogera have a similar density of primary and secondary schools. Diga has a lower density but nonetheless has the highest literacy ratio.

3.2.11 Livestock

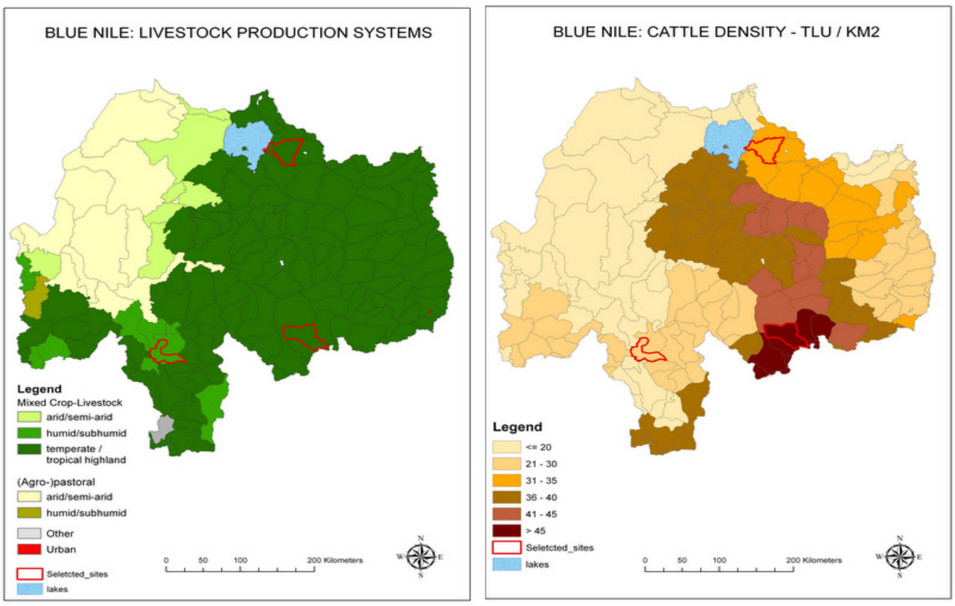

Figure 13 : Livestock production systems (left) and cattle density (right)

Figure 13 suggests that the mixed crop-livestock system have higher livestock densities than the (agro-) pastoral areas within the Blue Nile basin. All three study areas are mainly under mixed cropping and livestock keeping. Both Jeldu and Fogera are in the temperate/tropical highland agro-ecological zone (AEZ). Diga is located in the humid/sub-humid AEZ.

The areas that are located near to Addis seem to have the highest cattle density. From the three study landscapes Jeldu has the highest cattle densities. Around Lake Tana the livestock density is lower but still higher than in the Diga Woreda.

3.2.12 Crop yield and utilization

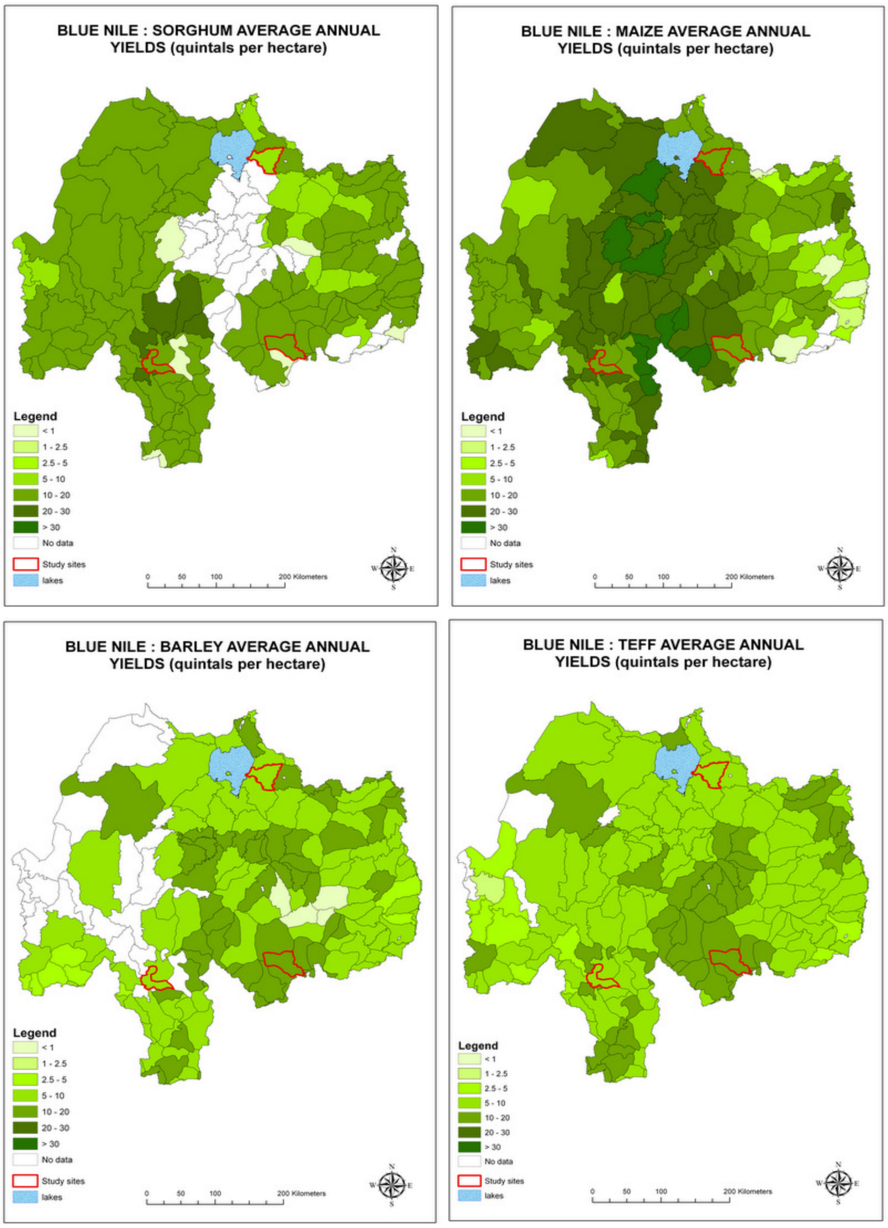

Figure 14 Crop yields for sorghum (top left), barley (top right), maize (bottom left) and teff (bottom right)

Figure 14 shows the different yield (quintal per ha) of the major crops, namely sorghum, barely, maize and teff. Sorghum is mainly grown in the dry areas, while maize in most productive in the humid areas. Teff and barley have their highest yield in the higher elevation with medium rainfall

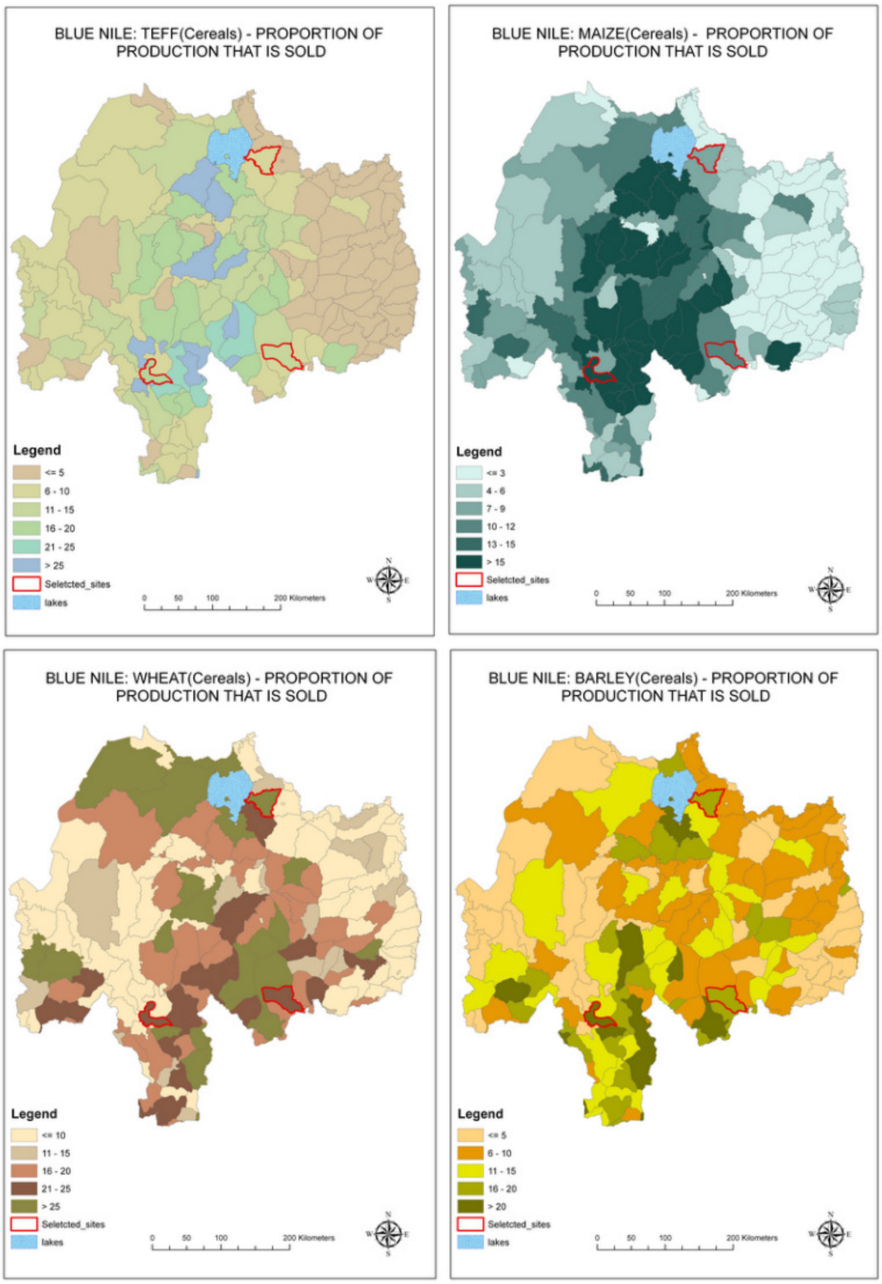

Figure 15 : Proportion of crop production that is sold : Teff (left), Maize (middle left), wheat (middle right), barley (right)

Figure 15 suggest that for all major crops, the proportion sold on the market is higher in the temperate humid areas with longer growing period, where also yields are generally higher. In Fogera and Jeldu a relatively high proportion of wheat and barley is sold, while teff and maize are not sold. In Diga, an important share of all the cereals is sold on the market.

3.2.13 Population density

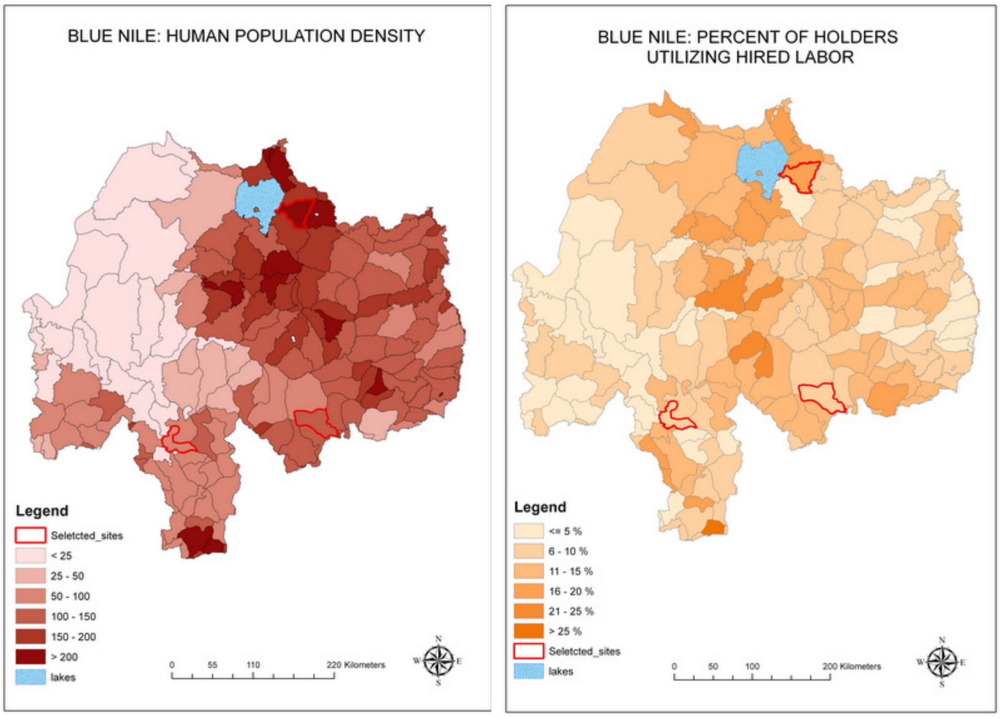

Figure 16 : Population density in the blue Nile basin

Figure 16 shows that population density is the highest around Lake Tana as well as on the highland plateau. Consequently Fogera is the most densely populated study site, followed by Jeldu and then Diga. The dry lowlands are the least populated areas. Also in terms of percentage of herd labor used, the Fogera study site differs from the two others, with more herd labor used in this Woreda.

3.2.14 Household composition

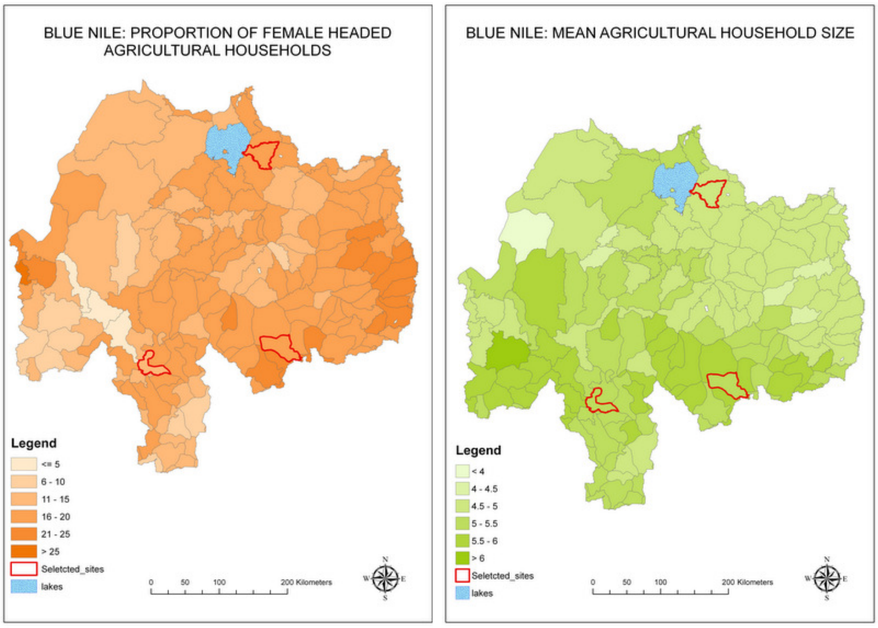

Figure 17 household composition: proportion of female headed households (left) and average household size (right)

Figure 17 shows that female headed households and small household size seem to be correlated, mainly in the lowlands and the mountainous areas. These are the less productive areas, where men migrate in the hope to find a job outside of their villages.

3.2.15 Credit and advise services

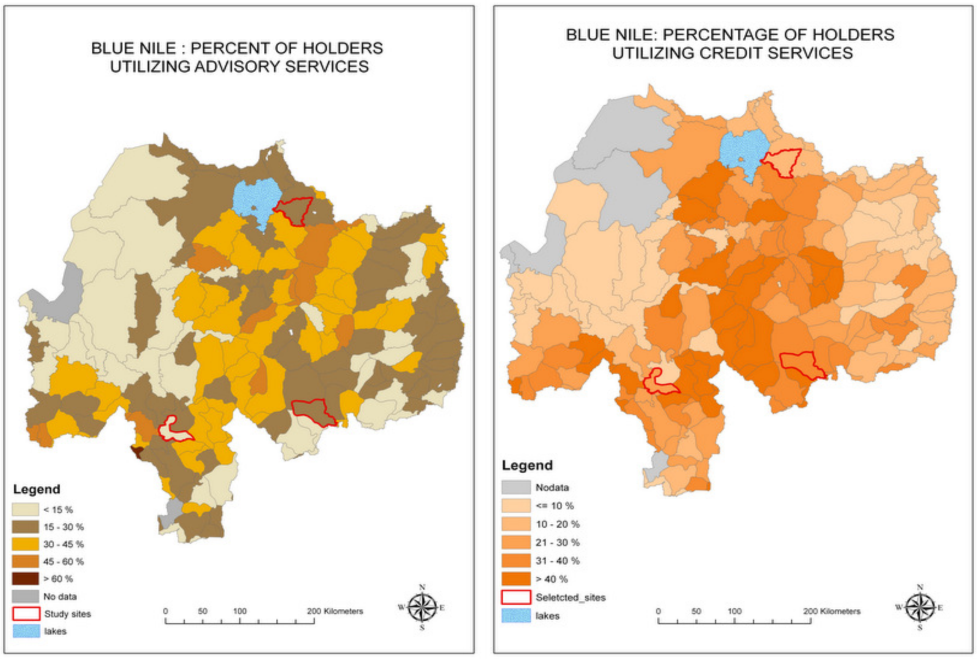

Figure 18 : Utilization of advisory (left) and credit (right) services

Figure 18 suggests the utilization of advices services and credit services show similar spatial patterns. Furthermore advice and credit services are mostly used in the highland plateau where crop yield is generally higher. Interestingly, these are not the locations that have the best market access, such as Fogera, where there is a low utilization of both advisory and credit services In Diga there is a very low utilization of advisory services. Jeldu exhibits a slightly higher utilization of credit services than the other two study sites.

3.2.16 Land holding

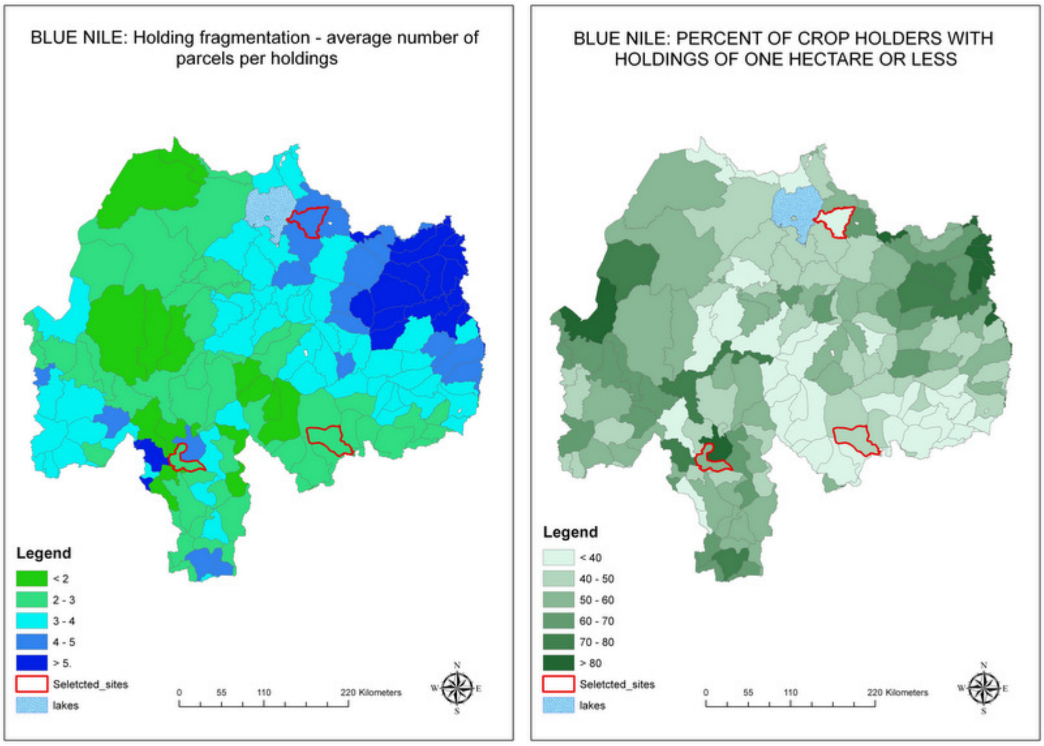

Figure 19 : land holdings : land fragmentation (left), landholding size low one ha (middle), share of rented land (right)

Land fragmentation is high on locations where landholding is small (below one hectare). These areas also are location where population density is high and therefore land is scarce. In locations where land is scarce, also the share of rented land is higher. Consequently Fogera with the highest population density is most fragmented with generally small landholdings and a high share of rented land .Diga and Jeldu both have relatively bigger landholdings and are not so much fragmented.

3.2.17 Income

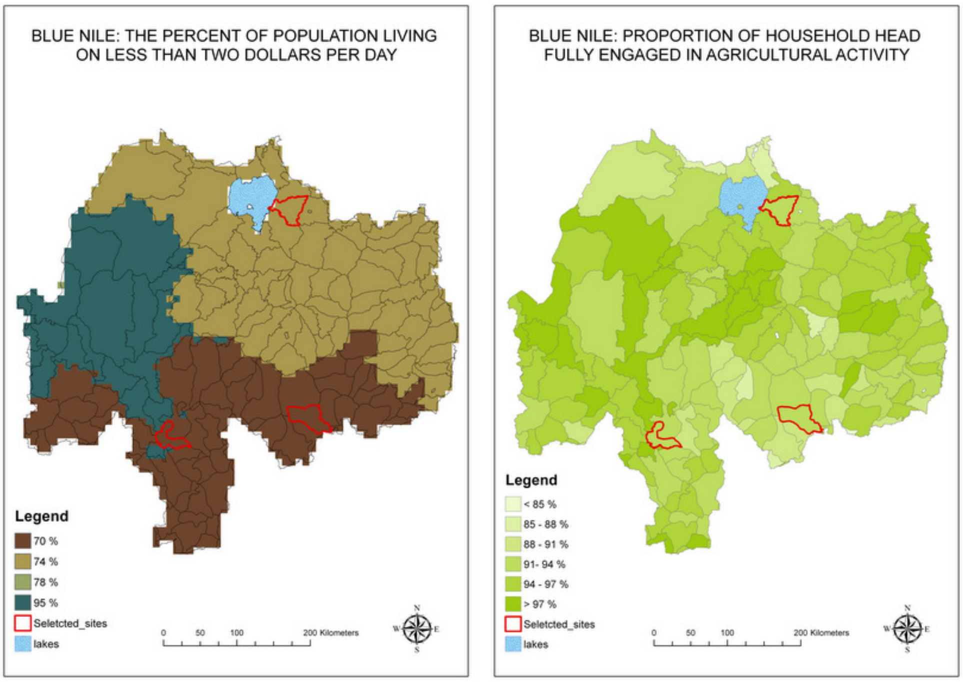

Figure 20 Income distribution: percent of population below the poverty line (2$ a day) (left), and the proportion of household heads fully engaged in agricultural activity (right)

Figure 20 suggests that the lowlands in the West are the poorest, while the South is the richest. The areas with high a proportion of households solely dependent on agriculture show two different patterns. First the poorest households in the West, do not have many off farm opportunities and mainly depend on agriculture. In the humid highland, where the length of the growing season is longer, yields are good and the population solely dependent on agriculture is less poor. While in the South, there are more off farm opportunities – probably also due to the proximity to Addis- and therefore dependency on agricultural activities is lower. The proportion of the population below the poverty line in this area is also lower.

4 Results for the set of characteristics

4.1 Factor analysis

Individually assessed characteristics in the previous section have been introduced into a factor analysis. Time consistency needs to be considered, because the data used has been collected from various geo-databases, namely ILRI, IWMI, IFPRI and FAO. Socio-economic characteristics are typically changing faster than bio-physical ones. It is therefore essential to only use data from similar years in order to avoid correlations that do not make sense. Therefore in this report two different factor analyses were carried out; the first one on socio-economic and institutional data, then a second on the rest of the variables. All the data used was collected in 2005 for the Ethiopian Rural Economy Atlas (2006), ensuring time consistency in the identified non-land based factors. Land-based variables change less quickly and therefore time consistency between the different variables is less of an issue.

A factor analysis has been run on biophysical and infrastructure and on socio-economic and institutional data separately. In order to identify the dimension of a factor and name it, two criteria have been used: a loading above 0.5 and at least two variables.

Table 3 : Factor loadings for biophysical and infrastructure variables

Variables used	Factor 1.1 *Drought potential*	Factor 1.2 *School density*	Factor 1.3 *Remoteness*	Factor 1.4 *Rainfall erosion potential*
Elevation	**-0.9717**	0.0784	-0.0338	0.0216
Minimum temperature	**0.9757**	-0.044	-0.0402	0.0276
Maximum temperature	**0.9712**	-0.0647	0.0401	-0.0006
Slope	**-0.5344**	-0.0572	**0.5743**	**0.2735**
Rainfall	-0.0085	-0.0181	-0.0401	**0.9333**
Coefficient of variation of rainfall	**0.5766**	-0.2063	-0.1022	**0.3821**
Travelling time to market	-0.0989	-0.0726	**0.8498**	-0.14
Road density	-0.3456	0.1299	**-0.7484**	-0.0098
Primary school density	-0.0986	**0.9704**	-0.0728	-0.0181
Secondary school density	-0.0432	**0.9782**	-0.0563	-0.0208

Table 3 shows the four factors loadings for land-based variables, in bold the one used to identify the dimension captured by each factor. The factor *drought potential* identifies locations on low elevation, with high temperatures, without steep slopes and a high variability of rainfall. The factor *school density* encompasses the density of primary and secondary schools. The factor remoteness indicates location with low road densities and long travelling time to major cities. These are also the more hilly areas, as suggested by the importance of the slope variable. Finally the factor *rainfall erosion* encompasses rainfall, variation of rainfall and slope. A high factor indicates location with high rainfall, high variability of rainfall and important slopes. These are areas where rainfall is likely to provoke important erosion. On the other extreme a low factor indicates locations with low rainfall, with low variability in flat area. These are areas where rainfall is less likely to provoke erosion. An intermediary factor results from low rainfall with high variability or high rainfall with low variability in hilly area, which are areas in which rainfall will provoke moderate erosion.

The factor analysis of socio-economic and institutional variable resulted in 5 factors, which loadings are shown in Table 4.

The first factor captures *dependency on agricultural production*. It suggests that smallholders with less than one hectare land fragmented on different parcels depend on agriculture only and do not hire labor. The second factor captures *demography* and related processes. It suggests that locations where many people live above the poverty line (2$) a day, are densely populated. In these areas land is scarce resulting in a rental market for land, explaining the importance of the

share of rented land in this factor. The third factor captures the *institutional* dimension and indicates locations where smallholders make use of credit and advice services. The fourth factor captures *household composition* dimension related variables and indicates locations with female headed households that have a small household size. Finally, the last factor captures *off farm income*. These are locations where smallholder not solely depend on agriculture but have some off farm income. This dimension also includes cattle population, suggesting that livestock is a form of investment when off- farm income is available.

Table 4 Factor loadings for socio-economic and institutional variables

Variables used	Factor 2.1 *Agricultural dependency*	Factor 2.2 *demography*	Factor 2.3 *institutions*	Factor 2.4 *household composition*	Factor 2.5 *off farm income*
Proportion of household heads fully engaged in agricultural activity only	**0.4952**	-0.2524	0.054	-0.0904	**-0.577**
Percentage of holders utilizing hired labor	**-0.6556**	0.3824	0.113	0.1818	-0.2035
Holding fragmentation - average number of parcels per holdings	**0.6924**	0.3805	0.0867	0.3413	-0.1792
Proportion of land area under smallholders that is rented	-0.1343	**0.6042**	0.4531	-0.0333	0.0211
Percent of crop holders who have land holding size of 1 hectare or less	**0.8841**	-0.0426	-0.1564	0.0614	-0.0062
Percentage of holders utilizing credit services	-0.0262	0.2372	**0.8714**	-0.0342	0.0434
Percentage of holders utilizing advisory services	-0.0952	-0.0324	**0.8818**	0.0417	0.0194
Proportion of female headed agricultural households	-0.0772	0.0697	-0.0446	**0.8255**	0.2686
Percentage of population living below the poverty line (2$)	0.128	**-0.8259**	-0.0576	0.1681	-0.2895
Average household size	-0.3045	0.0734	-0.0579	**-0.72**	0.3581
Population density	0.1272	**0.7941**	0.1082	0.2455	-0.1164
Density of cattle population	0.0902	0.0315	0.1032	0.0276	**0.8174**

4.2 Maps of the aggregate analysis of similarity

Each factor has been predicted for each woreda. The predicted factor does not have any meaning at such, but can be used to compare woredas among each other. In order to assess the spatial distribution of each factor, they have been mapped as shown in Figure 21. The Jenk's classification technique was used to define categories within which woredas are similar.

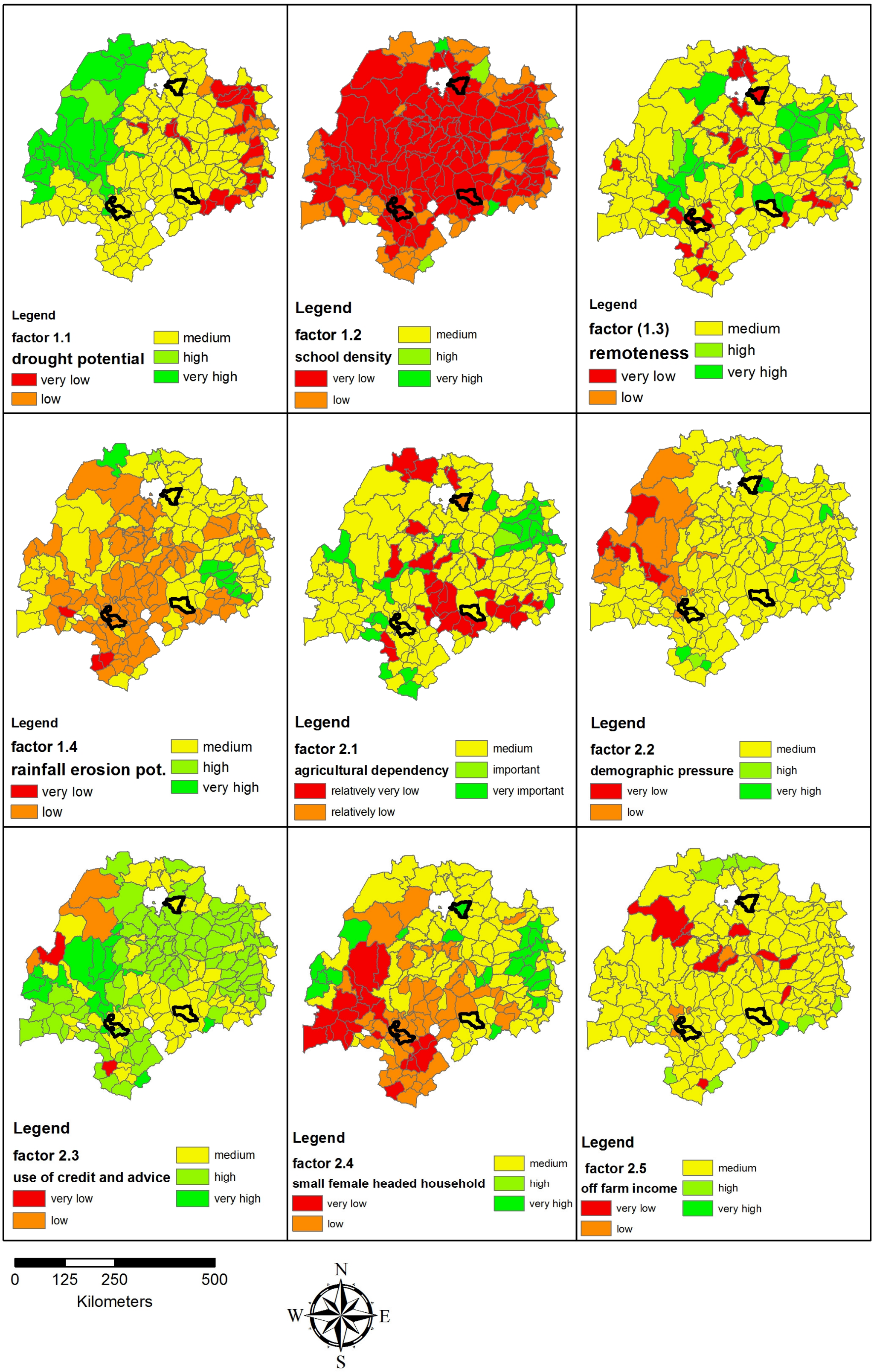

Figure 21 : Spatial patterns of the different identified factors, classified with a Jenk's criteria

Figure 21 shows the different spatial patterns of each of the previously identified factors. The drought potential factor (1.1) is mainly driven by elevation and identifies the lowlands as the most drought prone areas. All three study areas have a similar potential for drought. The access to education factor (1.2) shows that most of the basin has a poor access to education, with the exception of the area near to Addis Ababa, the area that borders the basin, which are the areas that have market proximity. Also in terms of education access all the three study sites have a similarly low access to education. The remoteness factor (1.3) shows that area around lake Tana but also in the South have a good accessibility while the rest of the area especially in the lowlands and the very mountainous areas are remote. This factor should be taken with cautions as roads are being built and urbanization is growing. The map indicated the Jeldu is the least accessible study site, while because of the newly built road, it is nowadays the most accessible of the 3 study sites. The factor rainwater erosion potential (1.4) shows that mainly the humid highlands are prone to rainfall erosion. Fogera and Jeldu being located in a more humid zone than Diga also have a higher potential for rainfall erosion.

The dependency of agriculture factor (2.1) shows that the central areas and areas near to Addis and Lake Tana depend less on agriculture than other areas. This can be explained by off-farm opportunities near to Addis, and the commercial farming and fishing around Lake Tana. In this perspective Fogera relies less on smallholder agriculture than the other two study site. The factor about demographic pressure (2.2) shows that demography and its consequences on landholding and wealth is pretty similar across the basin with the exception of the lowlands where population density is low and smallholders poor. The advice and credit services factor (2.3) shows that credit and advice is mainly used in the eastern part of the basin, and very few of this services are available for the lowlands. All three study sites have a medium use of credit and advice services. The household composition factor (2.4) shows that the remote areas (lowlands and mountainous regions) have the highest amount of small female headed households. Interestingly, Forgera also has a high amount of small female headed household. The last factor (2.5) reflects off-farm income, while it is also related to livestock ownership. Locations close to Addis have the highest off-farm income and also highest livestock ownership. All the three study sites have a similar off-farm income and livestock ownership.

5 Discussion

5.1 Similarity analysis

The classification and mapping of individual characteristic show the spatial variation of variables that have been identified as potentially important for RMS. Correlation between different variables can only be guessed. Factor analysis on the contrary is an exploratory data-driven approach that aggregates variables to form dimensions within which the variables vary in a similar way. It assumes an unobserved construct that generates the observed data. It allows to describe different processes and consequently the interactions leading to the observed patterns. The combination of both the individual analysis and the factor analysis allows to identify areas with similar characteristics: the lowlands in the West are drought prone, have little population and are poor. The areas around Lake Tana are relatively accessible; therefore

have access to off-farm opportunities also thanks to commercial farming. The very mountainous areas in the East are more remote, are more prone to rainfall erosion, have fewer infrastructures and have a bigger out migration of man resulting in a high number of female headed households. In the central part of the basin, where length of growing season is long, also advice and credit services are more intensively used.

The similarity analysis suggests that in general the three study sites are very similar. They mainly differ in terms of soil types, livestock identity, population and household composition. Fogera is the most populated area with the most small female headed households. It has a medium livestock density and a soil that needs to be well managed to be productive. Jeldu is the most livestock dense area on the most erosion prone soils, while Diga is the least densely populated area (due to only recent in-migration) with the lowest livestock density. It is located on relatively fertile soil and has the least amount of small female headed households.

Jeldu and Fogera are similar to each other when it comes to farming characteristics and infrastructure, such as length of growing period, density of schools, livestock holding or rainfall erosion potential. In contrast, Jeldu is much more similar to Diga in terms of socio-economics and institutional aspects, such as land fragmentation, wealth or use of credit and advice services.

This important similarity between the different study sites is surprising. Three reasons can explain this similarity : i factor analysis yields to wrong results because of missing variables, ii the scale chosen does not capture a sufficient level of detail, iii. the study site selection was suboptimal in terms of variability.

In order to test the hypothesis of missing drivers, one can compare other studies that use more expert based knowledge to assess the human-nature interactions. Both farming-systems, which reflect how farmers makes use of their land, and land-use which reflects how the whole society makes use of its land, given the bio-physical, socio-economic and institutional settings, can be used to assess variability. Figure 22 shows the farming systems as well as the land use map for the whole Blue Nile basin. Also according to these more expert based approaches, all three study sites are surprisingly similar. In all three, single cereal production is observed, however the major cereal varies: Diga is maize based, Jeldu and Fogera are teff based, part of Jeldu is barley based. The land use map also suggests that both Jeldu and Fogera are mainly used for cultivation: Fogera is more intensely cultivated than Jeldu. Land use in Diga is slightly different and indicates crop plantation which refers to the mango plantation. There are usually intercropped with maize explaining the maize based farming-system. Both the land use maps and the farming system map lead to a similar conclusion: all three study sites are quite similar suggesting that thought some variables could not be collected and included in this study, no major process explaining the surprising result has been missed in the factor analysis.

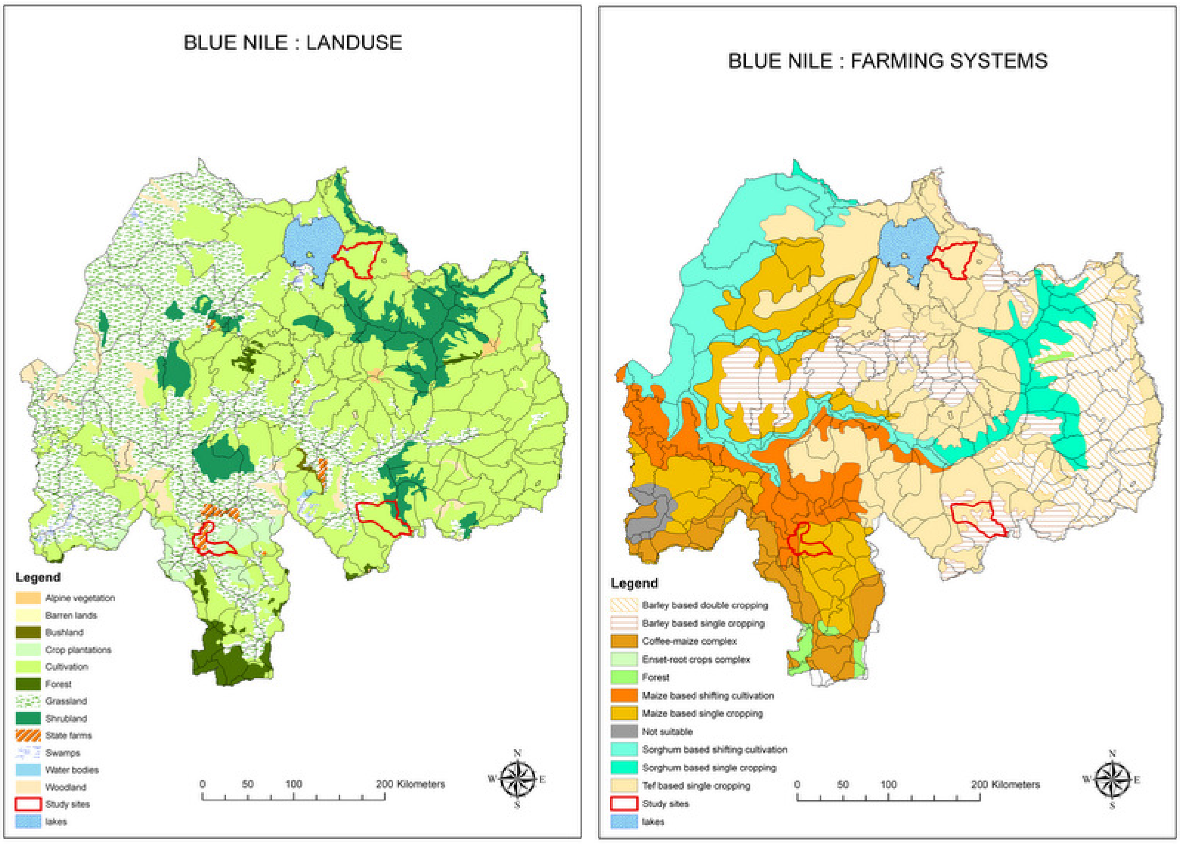

Figure 22 : Land use (left) and farming systems (right) source : ENTRO

Therefore, the similarity could be the result of the scale at which the similarity analysis has been run. The factor analysis has been run at basin scale making use of data at woreda level. The choice of this scale impacts our results in two ways: the extent of the data we use defines the heterogeneity that is introduced into the analysis and the resolution (woreda level) defines the level of detail by which this heterogeneity can be captured.

Firstly, the Blue Nile basin boundary has been used as extent of this study, including locations in very different agro-ecologies than the 3 study-sites, thereby increasing the overall heterogeneity within the analysis. This might increase the range of the analysis and therefore seemingly reduce the heterogeneity with the agro-ecologies of interest. To test this hypothesis, a factor analysis has been performed on a sub-set of woredas that have similar agro-ecologies than the 3 study sites. Results on the subset show the same characteristics than the basin wide analysis: the 3 study sites are quite similar. This suggests that the results are robust and are not dependent on the extent of the study.

Secondly, the resolution of the data was chosen at woreda level, the smallest level at which socio-economic data is available basin wide. Thus in order to model the basin extent, the heterogeneity within the woreda is lost. Indeed, there is a trade-off between the extent of the analysis and the level of heterogeneity that can be captured. For this reason, drivers to human-nature interaction may differ depending on the scale of the analysis (Cash 2006). Whereas at basin scale household and bio-physical

characteristics can only be taken into account on an aggregate level, human interactions, governance such as land redistribution or market orientation, location specific natural phenomena such as erosion or termites are lost. Therefore, when analyzed at woreda or micro-catchment scale, the three study sites might be very different from each other. To assess this scale, different data is needed from the farm as well as community and Kebele (municipality) level. A recent study identifies livelihoods zones in Ethiopia by integrating expert knowledge from community level (USAID). These zones are shown in Figure 23 and represent areas within which households on average share a similar livelihoods pattern, i.e. they have the same set of food, cash income sources and the same markets. The three study sites are all part of very different livelihood zones. Consequently, when using smaller scales and a higher level of detail, the three study sites seem to be very different. In this perspective, local action research and base-line assessment in the three sites will be important in order to identify dimensions within which the study sites differ which cannot be captured by the currently available data.

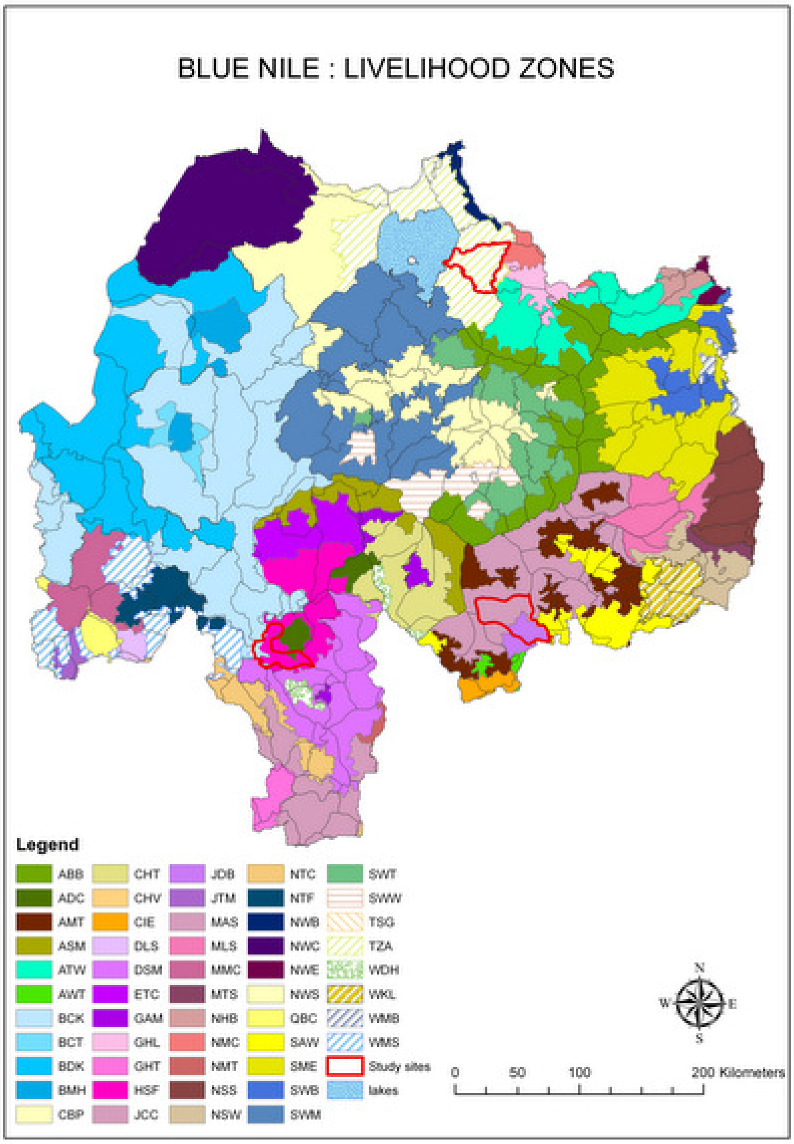

ABB	Abay Beshilo River Basin	GHT	Gera-Setema-Sale Forest, Teff. Honey & Cattle	QBC	Keto-Begi Cereals& Cattle
ADC	Abe Dongoro-Diga-Sasiga Coffee & Sorghum	HSF	Anger Maize, Sorghum & Finger Millet	SAW	Selale-Ambo Highalnd Barley, Wheat and Horsebean Belt
AMT	Ambo Selale Ginderbert Teff&Wheat	JCC	Jimma_illubabur Coffee, Ceral & Chad	SME	South Wollo Meher
ASM	Abay Gorge Sesame, Maize & Honey	JDB	Jaldu-Deni-Ilfata Barley & Potato	SWB	South Wollo Belg
ATW	Abay Tekeze Watershed	ITM	Illu_Wellega-Birbir Maize, Pepper & sesame	SWM	South West Maize, Finger Millet and Teff productive
AWT	Aneya-Wolisso-Ambo Teff & Cattle	MAS	Muger-Abay_jema Sorghum Teff Beld	SWT	South East Woyna Dega Teff
BCE	Bench Keffa Cereal Enset	MLS	Merhabete Lowland Sorghum and Teff	SWW	South West Woyna Dega Wheat
BCR	Basketo-Melo Coffee and Root crop	MMC	Mendi-Dabisu Maize, Sesame & Cattle	TSG	Tekeze Lowlnad Sorghum and goats
BDK	?	MTS	Mijnar Teff and Sorghum	TZA	Tana Zuria
BMH	?	NHB	North Wollo Highland Belg	WHO	
CBP	Central Highland Barley & Patato	NMC	North Easet Woyna Dega Mixed Cereal	WKL	Wuchale-Abischu-Kembebi Livestock, Wild Oates & Barley/Wheat
CHT	?	NMT	Nadda-Dilgel Gibe maize, Teff and Sorghum	WMB	South Wollo Meher and Belg
CHV	Cheffa Valley	NSS	North Shewa Highland Sheep and Barley	WMS	Wellega Coffee, Maize & Sorghum
CIE	?	NSW	North Shewa Highland Wheat and Teff Productive		
DLS	Dale-Lalo Sorghum & Maize	NTC	?		
DSM	Didess-Gibe Wama Valley Sorghum, Maize & Oilcrop	NTE	North Wollo East Plain		
ETC	Ebantu-Limu Teff and Cattle	NWB	N Highland Wheat Barley & Sheep		
GAM	Guduru-Amuru Maize, Teff and Cattle	NWC	North West Cash Crop		
GHL	Guna Highland	NWS	North West Sorghum Belt		

Figure 23 Livelihood zones (source US aid)

The comparison between a similarity analysis at basin scale and at woreda scale can give useful insights on cross-scale dynamics, which can then be taken into account for targeting and out-scaling interventions.

Finally it could be that the site selection is sub-optimal in terms of variability. Two hypotheses can be formulated to explain why these sites have been selected. Firstly, the selection was made by experts based on hydrological characteristics and the availability of partners to work. Experts might have looked at some of the characteristics separately without taking into account that many of these characteristics are correlated and that once this correlation is captured for example with a factor analysis, these sites are unexpectedly similar. Secondly, given the nature of the project focusing on action research, it might be that the availability of partners to work with is a criterion that has overweighed all the other criteria.

5.2 Targeting and out-scaling

Understanding the spatial heterogeneity of both bio-physical and socio-economic dimensions is a first step towards targeting and out-scaling interventions. Indeed, a successful intervention in one location can be a best-bet intervention in a location with similar characteristics.

This similarity analysis includes a multitude of possible drivers for adoption of any water related intervention. It does not consider that some of the interventions might be driven by different processes and therefore not all the variables used in this study might be relevant. Therefore, two major further steps are needed to develop an efficient tool for identifying best bet interventions: i. creating intervention/practice specific maps taking cross-scale dynamics into account and ii. Include a feedback loop with impact assessments and develop recommendation domain maps.

Firstly, the drivers of adoption and dis-adoption of rainwater management practices need to be understood. Suitability maps that include socio-economic and institutional dimension indicate those locations where all drivers suggest a suitable location for adoption of a practice. Methods for aggregation of different drivers into a suitability maps include principal component analysis (PCA) and multi-criteria analysis (MCA).

Clearly, the previously discussed cross-scale dynamics suggests that crucial drivers influencing farmers' decision making might not be captured at woreda level and consequently not be taken into account when suitability maps are built based on the data used in this similarity analysis only. In this perspective, expert knowledge about drivers identified at lower scale and up-scaling of farm household surveys are promising techniques to take dynamics from lower scales into account

Over the recent years, Ethiopia has been developing quickly, raising the issue whether data collected over the last decennia is relevant for creating similarity maps for current analysis and targeting. Indeed new roads are being built and need to be re-mapped as soon as new road maps are finished. In addition, not only increasing urbanization creates new markets for agricultural products, also new export channels are developing. This suggests that market access should be redefined to reflect current economic situation and accessibility taking into account the different actors in the value chain. Consequently, the factor remoteness should be remapped and should be reviewed regularly. Due to the increasing opportunities due to urbanization, also the livelihoods of smallholders might have changed around the growing cities and the newly build roads. Therefore, the maps presented in this report still need to be validated and – where possible- updated with recent field data[1], expert knowledge and action research. Only this will yield reliable maps indicating the current suitability.

It would also be interesting to take a forward looking perspective. The inclusion of projections of population density, precipitation, food and feed demand, market access, etc. could yield potential future suitability maps. These would be of special interest for strategic long-term planning.

After the development of suitability maps, a second remaining step involves taking into account of potential impacts. Based on the identification of suitable locations for specific interventions, a number of

[1] The baseline from N2 could be used to validate at least some of the maps

plausible scenarios can be built. These are in turn investigated in terms of their environmental and livelihoods impacts. The combination of bio-physical suitability, socio-economic adoption potential and expected impact then yields the final recommendation domains. Methods for aggregating and synthesizing these different components remains a challenge yet to be addressed, but could be along the lines of earlier work by Omamo et al. (2006a), Peden et al (2006c), Freeman et al (2008a) and Notenbaert et al.(2011).

6 Conclusion

This report aimed at developing a methodology that allows to identify different dimensions within which locations within a landscape differ. It makes use of a time consistent factor analysis of spatial data at the scale of the smallest unit at which data is available. This approach has been applied to the Blue Nile basin in the Ethiopian Highland. Factor analysis at woreda level identified four land based factors, namely drought potential, access to education, remoteness and rainfall erosion potential and five non-land based factors, namely agricultural dependency, demographical processes, institutions, household composition and off-farm income. Mapping these factors enables to identify spatial patterns that can be used for better targeting future research and interventions.

References

-, 2006. *Atlas of the Ethiopian Rural Economy*, International Food Policy Research Institute. Available at: http://www.ifpri.org/node/3763 [Accessed March 23, 2011].

Bai, Z.G. et al., 2008. *Global Assessment ofLand Degradation and Improvement1. Identification by remote sensing*, Wageningen, Netherlands: ISRIC World soil information. Available at: http://www.isric.org/isric/webdocs/docs/Report%202008_01_GLADA%20international_REV_Nov%202008.pdf.

Burel, F. & Baudry, J., 2004. *Landscape ecology concepts, methods, and applications*, Enfield, NH [u.a. Science Publ.

Cash, D.W., 2006. Scale and Cross-Scale Dynamics: Governance and Information in a Multilevel World. *Ecology and Society*, 11, p.article nr 8.

De Fraiture, C., Molden, D. & Wichelns, D., 2010. Investing in water for food, ecosystems, and livelihoods: An overview of the comprehensive assessment of water management in agriculture. *Agricultural Water Management*, 97(4), p.495–501.

Driessen, P.M. & Dudal, R., 1991. *The Major soils of the world: lecture notes on their geography, formation, properties and use*, Dept. of Soil Science, Agricultural University. Available at: http://books.google.com.et/books?id=MIWUAQAACAAJ.

Faures, J.-M. & Santini, G., 2008. *Water and the Rural Poor : Interventions for improving livelihoods in Sub-Saharan Africa*, Rome: FAO. Available at: ftp://ftp.fao.org/docrep/fao/010/i0132e/i0132e.pdf.

Freeman, A. et al., 2008a. Strategies and priorities for integrated agriculture for development in the SSA CP. Accra, Ghana, February 2008.

Hanjra, M.A., Ferede, T. & Gutta, D.G., 2009. Reducing poverty in sub-Saharan Africa through investments in water and other priorities. *Agricultural Water Management*, 96(7), pp.1062-1070. Available at: [Accessed 07:29:13].

Hanjra, M.A. & Gichuki, F., 2008. Investments in agricultural water management for poverty reduction in Africa: Case studies of Limpopo, Nile, and Volta river basins. *Natural Resources Forum*, 32(3), pp.185-202.

Humphreys, E. et al., 2008. *Improving rainwater productivity: Topic 1 synthesis paper.*, Colombo: CGIAR Challenge Program on Water and Food,.

Johnston, R. & McCartney, M., 2010. *Inventory of water storage types in the Blue Nile and Volta River Basins*, Colombo, Sri Lanka: IWMI.

Jones, P.G. & Thornton, P.K., 2000. MarkSim. *Agronomy Journal*, 92(3), p.445. Available at: [Accessed April 6, 2011].

Merrey, D.J. & Gebreselassie, T., 2010. *Lessons from Projects and Programs to Promote Improved Rainwater and Land Management in the Blue Nile (Abay) River Basin, Ethiopia*, CPWF.

Notenbaert, A., Herrero, M. & de Leeuw, J., 2011. *Dryland development options. In: Pastoralism and Biodiversity in Eastern Africa*, ILRI.

Omamo, S.W. et al., 2006b. *Strategic Priorities for Agricultural Development in Eastern and Central Africa*, International Food Policy Research Institute. Available at: http://www.ifpri.org/pubs/abstract/150/rr150.pdf [Accessed April 7, 2011].

Peden, D. et al., 2006c. Investment options for integrated water-livestockcropproduction in sub-Saharan Africa. *ILRI working paper*.

Rockström, J., Barron, J. & Fox, P., 2003. Water productivity in rain-fed agriculture: challenges and opportunities for smallholder farmers in drought-prone tropical agroecosystems. *Forthcoming. Water productivity in agriculture: limits and opportunities for improvement. Comprehensive Assessment of Water Management in Agriculture Series*, 1.

Rockström, J. et al., 2010. Managing water in rainfed agriculture--The need for a paradigm shift. *Agricultural Water Management*, 97(4), pp.543-550. Available at: [Accessed 09:06:12].

Schmidt, E. & Kedir, M., 2009. *Urbanization and Spatial Connectivity in Ethiopia:Urban Growth Analysis Using GIS*, IFPRI.

Snow, R.W. et al., 2005. The global distribution of clinical episodes of Plasmodium falciparum malaria. *Nature*, 434(7030), pp.214-217. Available at: [Accessed April 6, 2011].

Thornton, P. et al., 2002. *Mapping poverty and livestock in the developing world*, ILRI (aka ILCA and ILRAD).

USAID, *An Atlas of Ethiopian Livelihood : The Livelihoods Integration Unit*, Available at: http://www.feg-consulting.com/what/services/early_warning/livelihood-integration-unit-liu/liu-atlas/complete-atlas/Atlas%20Final%20Web%20Version%206_14.pdf.